人文社科类期刊长效文献研究

朱世琴 蒋辛未
华东理工大学科技信息研究所 上海 200237

摘要：[**目的/意义**]长效文献是指出版多年后仍然被连续多年引用，或者引用频次相对较高的文献，研究这部分老化较慢的长效文献的影响力，有助于完整评价文献整个生命周期的价值。[**方法/过程**]选取 CSSCI 数据库来源期刊(2016－2017)中九个学科在 2012 年－2016 年间出版期刊文献的引文年代数据，通过共时观察法进行科学文献老化情况和长效文献分析的研究。[**结果/结论**]期刊中引文年代较为久远的引用量来自于大量的低频次被引用的文献集合。长效文献数量较少，可分为 3 种类型：成长型、成熟型和衰退型，3 种类型的长效文献会随着时间推移产生一定转化。具有长效文献较多的期刊未必影响因子高，除心理学学科外，其他每个学科都有且仅有一本具有显著长效影响力的期刊。长效文献的关键词与当前学科研究热点相似，但是长效文献的追溯年代更久远，在反映当前学科热点来源的同时，对未来学科热点预测将发挥作用。

关键词：长效文献 文献老化 计量分析 CSSCI
分类号：G256
DOI：10. 13266/j. issn. 0252－3116. 2018. 19. 012

1 引言

自 1943 年 C. F. Gosnell[1]在其发表的博士论文中提出文献老化的概念，相关的研究已有 70 多年的历史，通过对文献老化规律的研究，产生了多种较为成熟的评价指标。传统的期刊文献评价主要是计算被引频次并结合一定的时间范围。其中较为经典的指标有 J. D. Bernal[2]借鉴物理学中物质半衰期的概念，提出“文献半衰期”；D. J. Price[3]在其研究中提出普莱斯指数；E. Garfield[4]提出影响因子并在 1975 年被 JCR 使用。从这些经典指标定义中的时间范围节点可以看出，这些影响力评价是基于较新文献的引用情况，计量指标的考察期相对较短，并且容易受到高被引文献的存量影响。因此计量指标也在不断发展，J. E. Hirsch[5]在 2005 年提出了 H 指数，L. Egghe[6]根据 H 指数提出了作者的影响力 G 指数。为了消除总被引频次的影响，俞立平[7]提出了历史影响因子。这些不断发展的评价指标，各有侧重，从不同的角度评价期刊，丰富了评价体系，但是存在局限于高被引文献，观察期较短，以及受历史数据存量影响等问题，对于出版时间较长的文献缺少相关研究与评价。

研究文献老化主要通过两个维度，即历时观察法和共时观察法。历时观察法通过对文献完整生命周期的连续观测来研究文献老化情况，对单篇文献的观察完整；共时观察法则是从引文的角度出发，有效地提升观察数据的准确性，谢海修和王宏鑫[8]利用灰色关联分析，对两者的差异进行分析，他们认为共时观察具有稳定性，适合某学科文献的老化研究。

生命周期理论应用于产品领域最早由 R. Vernon[9]提出，文献老化的过程可以视作文献从出版开始的生命周期，从以被引频次为主要指标的评价方式视角来看，文献的生命周期分为上升期、高峰期和下降期(见图 1)。每个时期的长短参照文献被引频次随着出版后年数的每年变动趋势而变化，分界点为引文峰值年代，在这之前的被引数量上涨时间段为上升期，引文峰值所在相邻的小段时间是高峰期，引文峰值出现后被引数量逐渐下降的时间段为下降期。国内魏瑞斌等[10]在国家社会科学基金项目“期刊论文创新性评价标准及其计量模型研究”中结合生命周期理论，提出被引生命周期的概念和被引比例的指标。之后，张中文等[11]结合生存分析，利用 Kaplan-Meier 估计量对文献生存率进行了分析，对文献的生命周期提出论文“寿终”“复活”“生存被引次数”等概念。

作者简介：朱世琴(ORCID：0000－0001－5646－9287)，副主任，副研究馆员，硕士生导师，E-mail：shqzhu@ecust. edu. cn；蒋辛未(ORCID：0000－0001－8061－7972)，硕士研究生。
收稿日期：2018－02－06 修回日期：2018－04－24 本文起止页码：93－100 本文责任编辑：王传清

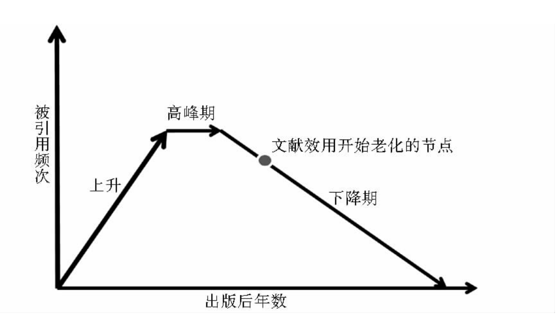


图 1 文献老化生命周期

目前文献老化注重老化规律、老化模型和各学科期刊的具体老化趋势研究，对于老化较慢文献的影响力评价研究较少，同时传统的文献指标评价倾向于关注文献高被引的情况。对于处于老化期的文献，其被引频次等影响力评价的关键指标往往会明显低于引文高峰期，但是结合其发文时间较长的特征，本研究认为也能够有效反映出其影响力，而且这种影响力是长效的。V. Larivière 等[12]通过对引文年代分布的研究验证了科研人员正在越来越依赖老文献。本研究借鉴生命周期理论并结合传统的文献计量方法，采用共时观察法，以 CSSCI 为数据来源，选取其中 9 个学科期刊在 2012 年 – 2016 年间发表的文献数据及其参考文献数据，整理各学科每年发表文献引文的年代分布数据，利用生存分析计算出期刊文献老化风险率开始出现上升的时间节点，根据确定的时间节点选取时间范围内的文献，进行频次、数量等指标的统计，从时间和频次两个维度筛选，定义这部分在老化阶段中仍然保持一定引用量的文献为长效文献并进行分析，以此来补充期刊文献的评价体系。

2 数据处理

本研究的数据来源于 CSSCI 数据库中来源期刊（2016 – 2017），选取其中 9 个学科的期刊文献被引数据，分别下载 2012 年 – 2016 年 5 个被引年的引文年代数据，数据下载时间为 2017 年 2 月。

考虑到以下两个原因，部分数据和期刊被剔除：①由于不同的期刊创刊时间不一致，无法满足本研究回溯年代的长度，以及部分期刊在数据库中的数据有缺失，为了保证数据的平稳性和有效性，部分不符合要求的期刊在数据下载中进行了剔除处理。②由于在数据下载的过程中，发现有极少量引用文献的年代早于 1975 年，出现的这部分文献集中于 20 世纪 30 年代或者建国后的初期，与后续出现引文次数的年代之间的间隔超过 10 年以上，这段时间刚好处于学术界的动荡时期，因而造成空档，对比期刊的创刊号时间，本研究将这部分引文数据也进行了剔除处理。因此，本研究利用共时观察回溯，以被引文献发表当年为基准年，设为第 0 年，以此基准来计算引用文献年代的长短，回溯最大范围确定为 40 年，最终被选取的学科和每个学科的期刊数量如表 1 所示：

表 1 选取学科与期刊数量

学科	法学	管理学	教育	经济	社会学	图书情报	心理学	哲学	政治学
选取期刊数（种）	19	27	26	69	10	20	6	11	30
收录期刊数（种）	21	29	36	73	10	20	7	12	32
筛选后文献数量（篇）	7 862	9 053	5 449	17 451	3 177	3 642	1 595	2 472	3 002
筛选前后比重（%）	97	94	93	95	89	93	93	97	89

本研究进行文献筛选时发现，CSSCI 数据库中的文献记录存在少量的错误。主要涉及 4 个方面：题目中标点符号错误、题目中错字漏字、应当含有小标题的题目缺失小标题，以及缺失作者。这些错误导致计算机在处理文献时，由于题目的差异，程序软件无法准确识别文献，将本应为一篇文献的引用切分为多篇，使得筛选文献的效率极大降低。切分多篇使得筛选的文献出现重复计数，同时，还将对应的引用量分开，使得引用量的统计不完整，而对后续长效文献的筛选结果产生影响。

3 长效文献研究

3.1 长效文献的定义

文献计量中涉及的文献概念，常见的有经典文献、高被引文献、零被引文献等，与本研究较为接近的是经典文献。王京山[13]对经典文献做出阐述，认为经典文献是兼具重要、规范、恒久特征的文献。本研究对象是在老化阶段中仍然保持一定引用量的文献，与“经典文献”通过定义比较，两者在文献时间跨度和价值上有相似之处，均关注时间长和效用高这两个方面；然而区别在于，相对于文献的整个生命周期，经典文献短期内效用很高，成为热点，随着研究深入，即使其研究观点可能会被更新的研究成果替代，其研究范式仍然具有经典性或者理论具有代表性，相比较，本研究的文献范围更大，因此尽管只有少量能在发表后成为热点，却能反映文献老化过程中的效用保持，可以作为期刊评价的补充，进一步完善期刊的评价体系。

因此本研究以文献老化为研究主题,对发表一定年份后仍然被多次引用的文献进行研究,结合生命周期理论,在时间和引用频次两个维度下,首次提出"长效文献"的概念。长效文献是指文献出版之后,经历相当长的一段时间仍然被连续多年引用,或者引用频次仍然相对较高的文献。

根据定义,长效文献需要一个精确的时间节点来确定文献出版后多长的一段时间可以视为时间长,同时也需要足够的连续多年引用。若既没有足够的引用频次,也没有连续多年引用,但是符合在时间节点之后仍然有少量被引的文献,称为久期文献。久期文献与长效文献是包含关系,长效文献⊂久期文献。

3.2 长效文献时间界定

长效文献的时间节点确定,本研究是利用生存分析中半参数估计的 Cox 回归方程(见公式(1))分析对文献老化情况的影响因素进行验证。

$$h(t) = h_0(t)\exp(\beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \cdots\beta_px_p)$$

公式(1)

本研究通过共时观察法,首先按年份分组,分别统计出每一种期刊在 2012 年－2016 年发表文献所引用文献的年代分布表,将文献发表的当年设定为第 0 年,依次向前观察期刊的引文年代分布。例如 2013 年发表的文献中引用了发表于 2000 年的文献,则该引用文献的年代记为第 13 年,而 2016 年发表的文献中引用了发表于 2000 年的文献,记为第 16 年。对每个组的第 N 年引用量进行累加计算,反映出每本期刊在 2012 年－2016 间的引文年代分布。以此分布结果计算累计引用频次、引文年代峰值、普莱斯指数和学科文献的半衰期,结合这些传统的文献计量指标以及期刊所属的学科作为影响因素放入模型,通过 Stata 软件进行计算。在回归分析的基础上,绘制人文社科类文献总体老化风险率,见图 2。从图 2 可以看出,从第 10 年开始,人文社科类文献的老化风险率曲线开始上升,直到第 30 年接近最高值,达到接近 0.6 的水平,即在引文年代的第 30 年还没有出现老化节点事件的期刊样本,在这一年中不再被引用的可能性达 60%。

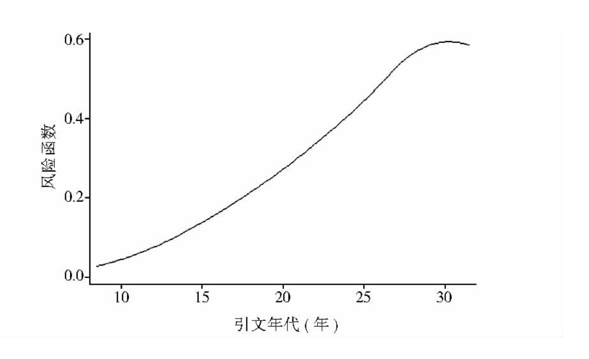


图 2 人文社科类文献总体老化风险率

3.3 长效文献效用界定

对于确定长效文献被引频次高低的界限问题,由于每个学科文献在被引数量、半衰期等多种因素的影响下,不仅存在数量级的差距,也会有指标极值的差异,因此单一地选择固定的界限是不合适的。根据筛选出的久期文献的被引频次分布,本研究针对不同的学科划定不同的引用频次的高低界限。将筛选出的久期文献按照被引频次从小到大进行编号排列,绘制散点图(见图 3),可以发现,所有学科都呈现出一致的规律,共时观察法下大量文献在发表 10 年之后作为引文出现的频次极低,均为 1 至 2 次,这部分文献的长期效用极小,而到达上升拐点时出现趋近 90 度的变化,变为直线上升。本研究的长效文献效用界定正是根据这部分拐点前后的平均引用频次计算得出。

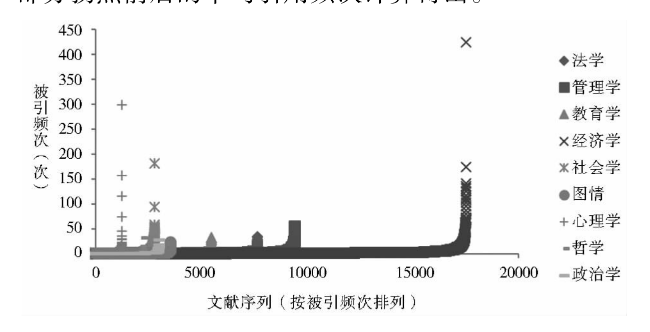


图 3 各学科久期文献的被引频次分布

同时,从图 3 还可以发现,每个学科的文献被引频次极值差异很大。心理学、社会学和经济学学科都可以发现离散于整体之外的单个的文献个体,这些文献在出版 10 年之后,相比于其他文献有着极高的引用量。通过对每个学科拐点附近的文献样本的被引量进行观察计算,确定每个学科的长效文献效用界限,结果如表 2 所示:

表 2 各学科效用界限

学科	法学	管理学	教育	经济	社会学	图书情报	心理学	哲学	政治学
被引频次(次)	10	10	9	10	10	8	10	8	8

3.4 长效文献筛选

根据上述长效文献的定义,在本研究共时观察的 5 年文献的引文年代中,首先确定长效文献为出版超过 10 年以上的文献,例如 2016 年发表的文献,其引文中能被视作长效文献的必须是于 2006 年之前发表,其次在出版 10 年之后仍能得到总计超过表 2 计算的所属学科效用的引用频次。

在时间长的前提下，本研究还设置了两个判定条件：①引用频次高，即文献在长效文献划定的时间范围内有足够高的引用量超过所属学科效用的引用频次；②连续引用，即在本研究的连续 5 年的观察周期中，一篇文献能够在不同的年份被作为引文引用，这种连续出现的引用能反映出文献所蕴含的价值具有延续性和前瞻性，是后续问题研究的始发点。结合这两个条件，本研究将长效文献分为 3 种类型：A 类文献同时满足上述两个条件，在观察期连续 5 年都有被引用，同时引用频次较高；B 类文献仅满足第二个条件，即在观察期内连续 5 年都有被引用，但引用频次不高；C 类文献仅满足第一个条件，文献引用的频次较高，但是并非连续 5 年都有被引用。

根据各学科效用界限以及是否在观察期内连续 5 年被引用这两个条件，本研究筛选得到各学科的长效文献（见表 3）。从表 3 可以看出，每个学科的长效文献占比不同，总体上占比非常少，占比最多的社会学和心理学学科有着 5% 左右的比重，占比较多的管理学和经济学学科，分别有 2.22% 和 3.53% 的占比，其他学科占比都低于 1% 。从 A、B、C3 类文献的数量来看，图书情报、哲学和政治学学科总量较少，同时，这 3 个学科的 B 类文献数量也较少，说明已经筛选出的 A 类文献基本就代表这些学科最近几年的主流趋势和研究方向。

表 3　各学科长效文献情况

学科	A 类文献（篇）	B 类文献（篇）	C 类文献（篇）	长效文献占比（%）
法学	27	19	32	0.99
管理学	66	29	106	2.22
教育学	14	16	12	0.77
经济学	220	55	342	3.53
社会学	74	20	79	5.41
图书情报	6	2	22	0.82
心理学	19	11	49	4.95
哲学	3	3	18	0.97
政治学	7	2	17	0.86

图 4 是对所有学科 A、B、C 3 种类型文献的年代分布情况箱线图。A 类和 B 类文献的出版年份中位数非常接近，都在 2000 年左右，C 类文献相较而言更为年轻，出版年份中位数在 2003 年左右。在出版年份的时间跨度上，无论是四分位的箱体宽度，还是边缘界限的宽度，A 类文献的跨度区间都明显大于 B 类和 C 类文献。对 A、B、C 3 种类型文献的上下四分位线和上下边缘线进行横向比较，可以看出，A 类上四分位线和下边缘线接近对应 B 类的上四分位和下边缘线，A 类的上边缘线接近对应 C 类的上边缘线，B 类的上边缘线和自身的上四分位线重合。这些特征反映出 A、B、C 3 种类型的长效文献随着时间变动会产生一定的变化。

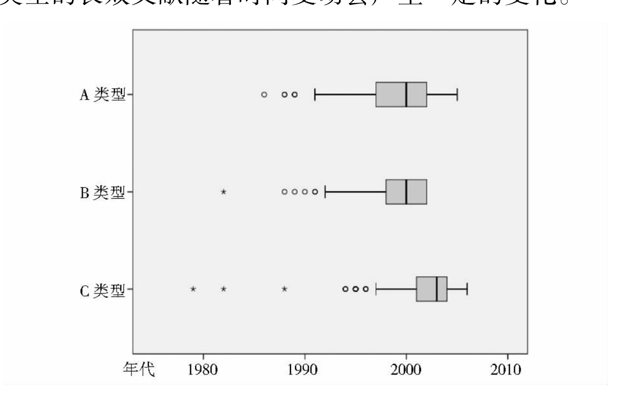


图 4　长效文献出版年代箱线图

3.5　长效文献各类型的转化

文献在进入老化期后没有被大量引用，但是保持少量被引频次的文献，在达到文献老化的判定时间节点（本研究统一设定为 10 年）之后变为久期文献，但随着时间的推移，连续多年被持续观察到被引，但引用量较低的变为 B 类文献，而部分仍然被大量引用，由于刚刚过 10 年，没有达到连续多年被观察到有引用的条件，转变为 C 类长效文献。C 类长效文献随着时间的推移，仍然能够保持大量的被引用频次便会升级为 A 类长效文献，不能保持的则不再是长效文献。A 类长效文献无法保持大量被引用频次的，但是仍然能够保持每年被引用的，则会变为 B 类长效文献，不能保持的则不再是长效文献。长效文献的转化趋势见图 5，综合这种变化的规律，A、B、C 3 种类型的长效文献可以分别对应为成熟型、衰老型和成长型文献。

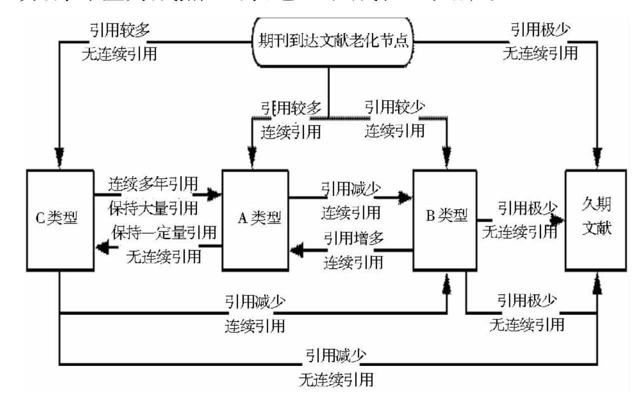


图 5　长效文献转化趋势

3.6　长效文献与影响因子关系

从含有长效文献的期刊占比来看，部分学科内的期刊影响力差异不大，但是将数量纳入统计之后，表 3 中

差异很小的管理学、社会学和心理学学科,学科内的期刊差异也变得非常明显。各学科内收录长效文献排名前5位的期刊见表4,其中心理学和哲学学科不足5种期刊。结合表4中每个学科排名第1位的期刊及其长效文献数量,与后面2－5位排名的期刊进行对比,除心理学之外的每个学科,排名第1位的期刊所收录的长效文献数量至少为排名第2位的期刊收录数量的两倍,其中管理学、社会学和哲学学科高出接近4倍,经济学学科更是高出8倍。这种巨大的数量差距说明除心理学之外的其他学科,尽管存在发文量不同的影响,但长效文献数量排名第1位的期刊在学科领域研究中占据着极其重要的位置和分量。心理学学科虽然排名第1的期刊没有独树一帜,但是将排名第2位的期刊也作为参照对象,排名第3位的期刊收录的长效文献数量与第一位相差3倍,与第2位相差两倍,说明心理学学科有着两种影响力较高的期刊,大量长效文献集中于这两本期刊之中,而其余的期刊收录情况与其他学科也是基本一致的。除图书情报学科外,学科内其他期刊收录的长效文献数量,排名第2、第3位的期刊相比排名第4、第5位的期刊,也有明显的数量差异,属于第二梯队。

表4　各学科收录长效文献排名前5位的期刊

学科	排名	期刊名	长效文献数量(篇)	期刊影响因子
法学	1	《法学研究》	33	3.26
	2	《中国法学》	17	4.88
	3	《法学》	6	1.84
	4	《现代法学》	4	3.25
	5	《中外法学》	4	2.05
管理学	1	《管理世界》	93	2.93
	2	《科研管理》	21	1.62
	3	《科学学研究》	15	1.58
	4	《系统工程理论与实践》	14	1.20
	5	《中国软科学》	11	2.29
教育学	1	《教育研究》	15	2.78
	2	《电化教育研究》	7	2.81
	3	《高等教育研究》	4	1.48
	4	《教育发展研究》	2	1.10
	5	《教育研究与实验》	2	0.53
经济学	1	《经济研究》	324	3.90
	2	《世界经济》	38	2.36
	3	《金融研究》	36	2.21
	4	《中国工业经济》	31	2.24
	5	《会计研究》	24	5.13
社会学	1	《社会学研究》	111	2.19
	2	《人口研究》	28	3.54
	3	《中国人口科学》	20	1.49
	4	《人口与经济》	9	1.61
	5	《妇女研究论丛》	3	0.48
图书情报学	1	《情报学报》	8	1.03
	2	《大学图书馆学报》	3	2.36
	3	《情报科学》	3	1.00
	4	《情报理论与实践》	3	1.05
	5	《图书馆》	3	1.11
心理学	1	《心理学报》	38	1.11
	2	《心理发展与教育》	28	1.17
	3	《心理科学进展》	9	1.16
	4	《中国临床心理学杂志》	5	1.19
哲学	1	《自然辩证法通讯》	11	0.30
	2	《哲学研究》	3	0.38
	3	《世界哲学》	1	0.19
政治学	1	《政治学研究》	11	0.90
	2	《世界经济与政治》	6	1.51
	3	《美国研究》	2	0.46
	4	《探索》	2	0.65
	5	《国际论坛》	1	0.42

影响因子是指该期刊近两年文献的平均被引用率,即该期刊前两年发表的论文在评价当年每篇论文被引用的平均次数。本研究参照维普数据库中的2015年各期刊影响因子,结合期刊的长效文献数量,发现两者并没有显著的相关性。所有学科中拥有长效文献最多的期刊其影响因子都不是最高的,其中管理学、教育学、图书情报学和政治学学科中影响因子最高的期刊都不在表4收录的长效文献排名前5位的期刊中。总体来看,一些影响因子较低的期刊也会存在部分长效文献。

这种情况说明,影响因子以期刊发表最近两年文献作为分析对象,其重点在于近两年的文献引用情况,这种影响力评价水平是基于较新文献的,但是对于久期文献的评价是不足的,忽视了久期文献所具备的长效影响力。

另一方面,影响因子是一个动态的评价指标,基于近两年文献,每年的影响因子会随着年份变动而变化。本研究的长效文献是基于共时观察法,所得数据同样也是一种时间截面的计量结果,不同年份的长效文献数量会发生变动,相比于历时观察法,能够避免完整观察全部引用量而造成的累积效应。因此将长效文献的数量作为评价指标引入期刊评价,能够更加全面地反映期刊影响力。

4 经济学科长效文献知识网络

由于长效文献的数量较少,本研究仅选取文献数量最多的经济学科,对所有的长效文献进行知识网络分析,由于 CSSCI 的来源查询无法获取 1998 年之前的文献详细信息,因此本研究采用共时观察法,由 CNKI 获取并筛选出 1976 年－2016 年间的 CSSCI 经济学类长效文献的基本信息,利用 CiteSpace 对这些文献的关键词进行共现分析,频次大于等于 3 次的关键词共现结果按年份排序见表 5。从表 5 中可见这些关键词来源文献并非分散于 1976 年－2016 年间,而是主要集中于 1999 年－2003 年间,期间包含了 25 个关键词。经济增长是出现频次最高的关键词,其次是上市公司、公司治理和人力资本,频次均超过 10 次。这段时期内长效文献的关键词反映出,经济学学科中,经济增长是研究的大背景,在该背景下,对于企业治理,尤其是上市公司的管理是研究热点。

2014 年,马费成等[14]对 1998 年－2011 年间的经济学文献进行了计量分析,其研究的年代与本研究长效文献的关键词共现结果集中的年代非常接近,对比其研究结果可以发现,按年份排序,本研究长效文献的共现关键词结果与马费成等通过 1998 年－2011 年期间所有 CSSCI 经济学类全部文献关键词共现得出的结果非常接近。马费成等研究得出的热点图,在 1998 年－2000 年阶段热点关键词是国企改革、上市公司等,而本研究长效文献的共现结果不仅发现了国有企业是研究热点,还发现节点事件提前到了 1995 年。同样,对于人力资本的研究,马费成等研究发现是在 2001 年－2003 年阶段,而本研究发现节点事件出现在 1996 年。尽管随着出版年份的久远,长效文献的被引频次在引用量上无法与引文高峰期相比,但是这些长效文献的关键词共现结果与全部文献的关键词共现结果非常接近,并且时间节点还有所提前。马费成等的研究是观察年份期间文献的关键词热点,本研究通过长效文献的关键词对比,反映出长效文献一定程度上是研究热点起始的年代,对未来学科研究热点具有指导性,可以作为当前学科研究热点参照的目标和未来热点的预测方式之一。

表 5 经济学长效文献关键词共词统计信息

序号	频次	中心性	年份	关键词	序号	频次	中心性	年份	关键词
1	46	0.43	1993	经济增长	20	6	0.03	2001	股权结构
2	4	0.01	1993	经济改革	21	4	0	2001	金融发展
3	5	0.04	1994	效用函数	22	14	0.02	2002	公司治理
4	6	0.04	1995	国有企业	23	6	0.04	2002	技术进步
5	5	0	1995	生产函数	24	6	0	2002	资本结构
6	3	0	1995	规模经济	25	5	0.11	2002	中国
7	11	0.06	1996	人力资本	26	4	0.02	2002	中国经济
8	5	0.02	1997	资本投入	27	4	0	2002	外商直接投资
9	5	0.01	1998	收入分配	28	4	0	2002	基尼系数
10	4	0	1998	对外直接投资	29	4	0.01	2002	公司绩效
11	5	0	1999	净资产收益率	30	4	0	2002	中小企业
12	4	0.14	1999	金融机构	31	6	0.02	2003	全要素生产率
13	4	0	1999	市场结构	32	4	0	2003	地区差距
14	3	0	1999	现金流量	33	3	0.01	2003	企业业绩
15	6	0.32	2000	股票市场	34	3	0	2003	资本存量
16	5	0.14	2000	影响因素	35	3	0	2003	产业结构
17	3	0.01	2000	经济发展	36	3	0	2004	城市化
18	11	0.12	2001	上市公司	37	3	0	2005	技术效率
19	9	0.19	2001	货币政策					

通常高频次的关键词被用来确定一个研究领域的热点,而陈超美[15]认为中心性超过 0.1 则具有较强的影响力,说明该词所起的承上启下作用越大,也即研究热点之间的转化关系越强,从而可探析出现有研究的知识逻辑演进脉络。参照经济学类长效文献关键词图谱以及表 5 中的中心性,可以发现一些关键词的聚类。经济增长是最大的大类,其下包括公司治理、上市公司、货币政策和股票市场等,说明经济学类研究的对

象。第二类是技术进步、效用函数和生产函数等经济学数学模型方面的研究。第三类是人力资本、净资产收益率等企业管理评价的具体内容。从表5中可以看出,中心性高的除了经济增长这个背景关键词之外,其他高频词的关键词未必中心性高,从年份来看中心性较高的关键词集中于2000年－2002年,刚好处于长效文献3种类型的年份的中位线附近。

利用CiteSpace软件进行聚类处理,结果见图6,圆圈越大字体越大,该关键词出现的频率越高,颜色接近的关键词之间有着较为紧密的关联。从颜色聚类的区分结果可以发现,经济学科的长效文献在经济增长的背景下,研究热点分为两个方面,即企业运营管理、市场政策资源等方面,以及管理技术、管理评价、评价计量指标和方式等内容。



图6　经济学类长效文献关键词图谱

5　结语

本研究借鉴生命周期理论,基于共时观察法抽选CSSCI中的9个学科,通过生存分析方法确定了长效文献的时间节点为10年,并参照连续被引年份和引用量两个维度筛选出各个学科的长效文献进行分析,研究发现:

(1)研究者对久期文献的引用范围非常广泛,少有集中性,大部分久期文献的被引用频次只有1－2次,期刊中较为久远引文年代的引用量来自于大量的低频次被引用的文献集合。

(2)本研究选取的期刊中,除心理学学科外,其他每个学科都有且仅有一本具有显著长效影响力的期刊,学科期刊的长效影响力呈现出一超多强,多层次的格局。

(3)符合本研究筛选条件的长效文献数量极少,长效文献根据被引频次和连续被引用年数两个维度被分为3种类型:成长型、成熟型和衰退型,这3种类型的长效文献,会随着时间推移产生一定转化,这种转化反映出期刊老化的规律。

(4)对长效文献数量较多的经济学学科利用CiteSpace软件绘制关键词共现图谱,研究发现长效文献能够作为评价出版文献相对久远的期刊的有效补充。同时,通过对经济学科全部文献关键词共现图谱的对比,发现这些长效文献的关键词与当前学科研究热点相似,但是长效文献的年代更久远,在反映当前学科热点来源的同时,也对未来的学科热点预测发挥作用。

本研究局限于CSSCI中随机抽取的9个学科,对理工类学科有待进一步研究。长效文献时间节点计算是基于本次选取数据回归所得,若进行单学科或其他年份的研究,10年的时间节点未必绝对准确。同时,共时半衰期的观察并不能涵盖文献老化过程的完整生命周期,并且会随着选取年代的变化而变化。但是这种动态的变化,与影响因子类似,因此长效文献的研究有助于进一步完善期刊文献影响力的评价,对文献整个生命周期的高峰之后部分进行补充评价,对文献价值的认知会更为完整。

参考文献:

[1] GOSNELL C F. The rate of obsolescence in college library book collections as determined by an analysis of the three select lists of books for college libraries[D]. New York: New York University, 1943.

[2] BERNAL J D. The transmission of scientific information: a user's analysis[C]// Proceedings of the international conference on scientific information. Washington: National academy of science, 1959: 85－95.

[3] PRICE D J. Citation measures of hard science, soft science, technology, and nonscience[M]// Nelson C E. Communication among scientists and engineers. Massachusetts: Heath Lexington Books, 1970: 3－22.

[4] GARFIELD E. Citation analysis as a tool in journal evaluation - journals can be ranked by frequency and impact of citations for science policy studies[J]. Science, 1972, 178(4060):471－479.

[5] HIRSCH J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569－16572.

[6] EGGHE L. Theory and practise of the g-index[J]. Scientometrics, 2006, 69(1): 131－152.

[7] 俞立平. 历史影响因子:一个新的学术期刊存量评价指标[J]. 图书情报工作,2015,59(2):89－92.

[8] 谢海修,王宏鑫. 文献老化的历时观察与共时观察的灰色关联分析[J]. 情报杂志, 1997,16(6):24－26.

[9] VERNON R. International investment and international trade in the product cycle[J]. The quarterly journal of economics, 1966, 80

(2): 190-207.
[10] 魏瑞斌,田大芳,武夷山. 基于被引生命周期的期刊被引的比较研究[J]. 图书情报工作,2014,58(1):79-85.
[11] 张中文,徐天和,董秀芬,等. 学术期刊论文生存被引次数的定义与应用[J]. 编辑学报,2015,27(4): 316-319.
[12] LARIVIÈRE V, ARCHAMBAULT E, GINGRAS Y. Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900-2004)[J]. Journal of the Association for Information Science and Technology, 2008, 59(2): 288-296.
[13] 王京山,王锦贵. 经典文献概念分析[J]. 图书与情报,2006(1):103-105, 115.
[14] 马费成,刘旻璇. 知识网络的结构、演化及热点探测——CSSCI(1998-2011)经济学文献计量分析[J]. 情报科学,2014,32(7):3-8.
[15] CHEN C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the Association for Information Science and Technology, 2006, 57(3): 359-377.

作者贡献说明:
朱世琴:提出研究思路与方法,设计论文框架,修改论文;
蒋辛未:处理与分析数据,撰写论文。

Analysis of Long-utility Literatures for Humanities & Social Sciences Journals

Zhu Shiqin Jiang Xinwei
Institute of Science & Technology Information, East China University of Science & Technology, Shanghai 200237

Abstract: [**Purpose/significance**] Long-utility literature refers to the literature still being cited for a long time with a relatively high frequency of citation after publication for many years. The influence of long-utility literature contributes to fully evaluate the entire life cycle of the literature. [**Method/process**] This paper studies literature obsolescence and long-utility literature analysis through synchronic observation based on the citation age data of literature which published from 2012 to 2016 in nine disciplines in CSSCI. [**Result/conclusion**] It is found that earlier citation in the journal came from a large number of low frequency references. There are only a few long-utility literatures which can be divided into three types, that is, growing type, mature type and recess type, and there is a certain conversion relationship between three types as time goes on. Journals with more long-utility literature may not have high impact factor. There is only one journal with significant long-utility influence in each discipline except psychology. The keywords of the long-utility literature are similar to those of current research hotspot, but the long-utility literature can be traced far more distant. It will play an effective role in reflecting the current hotspot and forecasting the trend in the future.

Keywords: long-utility literatures literature obsolescence bibliometric analysis CSSCI